Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States

David Mendez, Kenneth E. Warner, and Paul N. Courant

From 1965 to 1990, the prevalence of cigarette smoking among US adults (aged ≥18 years) fell steadily and substantially. Data for the 1990s suggest that the smoking initiation rate is increasing and that the decline in the prevalence of smoking may have stalled, raising the fear that the historical 25-year decline will not continue. The authors used a new dynamic forecasting model to show that although the decline may slow down, the demographics of smoking imply that prevalence will inexorably continue to decline over the next several decades, even without any intensified efforts aimed at tobacco control. The authors estimated and validated the model using historical (1965-1993) data collected by the National Health Interview Surveys on the prevalence of smoking among adults. Their results indicate that the current increase in the smoking initiation rate partially explains the fact that the prevalence of smoking has apparently leveled off, but even if the most grim assumptions about future initiation rates are used, the prevalence of smoking among adults will continue to decline for several more decades. The authors predict that if current initiation and cessation behaviors persist, the prevalence of smoking among adults will automatically decline from its current level of 25% to 15-16% by the second quarter of the next century. Even so, smoking will remain the nation's leading cause of premature death.

The prevalence of smoking among US adults fell steadily and substantially for a quarter of a century following the 1964 publication of the first Surgeon General's report on smoking and health (1). In 1965, the prevalence was 42 percent; by 1990, it had fallen to 25 percent (2). The decline was particularly notable among men (prevalence fell from 52 percent to 28 percent, respectively). However, recent data suggest that the decline may have stalled. Government surveys through 1994 show that the prevalence of smoking among adults is holding steady at about 25 percent (3). Data on the 30-day prevalence of smoking among high school and junior high school students, the harbinger of future smoking patterns, show annual increases since the beginning of the decade (4). Health officials have expressed alarm at the apparent leveling off of the prevalence of smoking among adults and the potential for a future increase, as suggested by the increasing number of teenagers who smoke.

We think that the alarm is misplaced. Although it is certainly possible that the decline in the prevalence of smoking will slow, the demographics of smoking imply that prevalence will inexorably continue to decline over the next several decades, even without any intensified efforts aimed at tobacco control. Indeed, if it is assumed that contemporary patterns of smoking initiation and cessation do not change, adult prevalence will automatically fall to about 16 percent by the end of the first quarter of the next century. This decline will be the inevitable result of the aging of the population and of anticipated birth and mortality patterns, combined with peak cohort-specific values for smoking prevalence, contemporary prevalence, and cessation rates. It can be understood by recognizing that today's higher prevalence of smoking reflects the smoking (and cessation) histories of many cohorts of smokers who joined the pool at different (and, over time, decreasing) initiation rates. For example, when the cohort of current (1997) 50-year-olds was aged 18 years, their prevalence of smoking was 46 percent. Such a high prevalence for people aged 18-24 years has not been observed since then (1965). Even the most grim predictions about future initiation rates are
far below 46 percent. Although initiation rates may be rising and may continue to rise in the near future, the overall prevalence of smoking is still likely to decline, because older smokers who joined the pool at extremely high prevalence rates continue to boost the aggregate rate of smoking. These smokers are gradually disappearing from the pool because of death and smoking cessation, both of which will continue to reduce the overall prevalence. In this paper, we describe the model used to derive this conclusion and consider how robust our findings are concerning plausible changes in smoking behavior among young people and in cessation behavior among older cohorts.

MATERIALS AND METHODS

Model

We used a dynamic model to forecast the future prevalence of smoking among US adults (aged ≥18 years), computing prevalence over time by tracking the inflow of new smokers and the outflow of smokers caused by death or smoking cessation (figure 1). The mathematical specification for the model follows:

\[
P_{a,t} = P_{a-1,t-1} \times (1 - \delta_{a-1,t-1}) [a = 1, ..., 110] \quad (1)
\]

\[
P_{0,t} = \alpha_t \quad (2)
\]

\[
C_{a,t} = C_{a-1,t-1} \times (1 - \mu_{a-1,t-1})
\times (1 - \beta_{a-1,t-1}) [a = 19, ..., 110] \quad (3)
\]

\[
C_{18,t} = \gamma_t \times P_{18,t} \quad (4)
\]

\[
\sum_{a=a_i}^{a_f} C_{a,t} \quad (5)
\]

\[
R_{(a_i;a_f),t} = \frac{\sum_{a=a_i}^{a_f} P_{a,t}}{\sum_{a=a_i}^{a_f} P_{a,t}} \quad (5)
\]

\[
\mu_{a,t} = \sum_{a=a_i}^{a_f} P_{a,t} \quad (6)
\]

where:

- \( P_{a,t} \) = size of the cohort aged \( a \) years in year \( t \)
- \( C_{a,t} \) = number of current smokers aged \( a \) years in year \( t \)
- \( R_{(a_i;a_f),t} \) = prevalence of smoking among individuals aged \( a_i \) to \( a_f \) years in year \( t \)
- \( \delta_{a,t} \) = death rate in year \( t \) among individuals in the general population aged \( a \) years
- \( \beta_{a,t} \) = death rate in year \( t \) for smokers aged \( a \) years
- \( \mu_{a,t} \) = smoking cessation rate in year \( t \) among smokers aged \( a \) years
- \( \gamma_t \) = prevalence of smoking among 18-year-olds in year \( t \)
- \( \alpha_t \) = size of the birth cohort in year \( t \)

The number of people aged \( a \) years in year \( t \) was computed by multiplying the number of people aged \( a - 1 \) years in year \( t - 1 \) by the appropriate survival rate (1 - death rate). Birth cohort sizes were supplied exogenously to the model. The model ignores the effect of migration on future population size and the prevalence of smoking. Patterns of smoking among future immigrants cannot be predicted. Given the small size of the immigration pool with respect to the domestic population, this omission is not likely to bias our results substantially.

Death rates were differentiated by year, age, and smoking status. The number of current smokers in any given year was estimated as the number of current smokers in the previous year who survived to the current year and did not stop smoking. The prevalence of smoking at age 18 years, supplied exogenously to the model for each cohort under study, was used to calculate the size of each year's cohort of new adult smokers. The prevalence of smoking among any specific age group in a specific year was computed as the ratio of current smokers to the total number of people in the group in that year.

The model assumes that no smoking initiation occurs after age 18 years, when prevalence attains its peak value for the cohort. After age 18 years, smoking cessation drives the dynamics of the model. This is consistent with data showing that by the 1980s, almost all regular smokers began smoking before age 20 years; by 1991, the mean age of becoming a daily smoker was 17.7 years (5).

The model uses smoking cessation rates as the parameters to be estimated from historical data on the prevalence of smoking. The model recognizes that cessation rates have changed historically with age and cohort. Existing data show that among cohorts born before 1935, cessation rates appear to be influenced
primarily by calendar year, whereas cessation rates for more recent cohorts appear to be determined by age; cessation rates are generally higher among those who are older (6). Preliminary analysis of the available data led to the age-group breakdown given in expression 6 above. Additionally, it was particularly important to allow for different values of the cessation rate before and after 1980. Given that our observations consisted of prevalences of smoking, we were only able to estimate net rates of change in prevalence; that is, initiation and cessation rates were necessarily confounded in our estimates. Prior to 1980, a nonconsequential proportion of individuals, particularly women, became regular smokers after age 20 years. Therefore, rates estimated for that period do not truly correspond to cessation rates, particularly among those in the younger age group. This was not the case after 1980, when people rarely began smoking after age 20 years.

Parameter estimation

To use the model to forecast the future prevalence of smoking, we first had to estimate the smoking cessation rates used by the model. The estimation procedure could also be used to test the conformance of the model to historical data. The procedure is described below.

Data. To estimate cessation rates, we collected data on the past prevalence of smoking. Specifically, we used data on the prevalence of smoking among groups aged 18–24, 25–44, 45–64, and ≥65 years in 1970, 1974, 1978–1980, 1983, 1985, 1987–1988, and 1990–1993. These prevalences were estimated by the Centers for Disease Control and Prevention from the National Health Interview Surveys (2). The data are shown in table 1.

The National Health Interview Survey is a principal source of information on the health of the civilian noninstitutionalized population. The annual survey consists of a basic set of health, socioeconomic, and demographic questions. To determine the prevalence of smoking among adults, defined as individuals aged 18 years or older, the survey collects self-reported information on cigarette smoking from that group. Prior to 1992, the survey defined current smokers as those who answered “yes” to the following two questions: “Have you smoked at least 100 cigarettes in your entire life?” and “Do you smoke cigarettes now?” Since 1992, the second question has been replaced by the following: “Do you smoke cigarettes every day, some days, or not at all?” Current smokers are now defined as those who have smoked at least 100 cigarettes and now smoke either every day or on some days. According to the Centers for Disease Control and Prevention, explicitly including “some days” of smoking in the definition of current smoking increases the overall estimate of the prevalence of smoking by

<table>
<thead>
<tr>
<th>Year</th>
<th>18–24</th>
<th>25–44</th>
<th>45–64</th>
<th>≥65</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>0.380</td>
<td>0.446</td>
<td>0.386</td>
<td>0.161</td>
<td>0.374</td>
</tr>
<tr>
<td>1974</td>
<td>0.376</td>
<td>0.445</td>
<td>0.377</td>
<td>0.173</td>
<td>0.371</td>
</tr>
<tr>
<td>1978</td>
<td>0.344</td>
<td>0.383</td>
<td>0.367</td>
<td>0.163</td>
<td>0.341</td>
</tr>
<tr>
<td>1979</td>
<td>0.344</td>
<td>0.389</td>
<td>0.348</td>
<td>0.164</td>
<td>0.335</td>
</tr>
<tr>
<td>1980</td>
<td>0.333</td>
<td>0.378</td>
<td>0.356</td>
<td>0.172</td>
<td>0.332</td>
</tr>
<tr>
<td>1983</td>
<td>0.342</td>
<td>0.363</td>
<td>0.333</td>
<td>0.167</td>
<td>0.321</td>
</tr>
<tr>
<td>1985</td>
<td>0.293</td>
<td>0.348</td>
<td>0.316</td>
<td>0.160</td>
<td>0.301</td>
</tr>
<tr>
<td>1987</td>
<td>0.271</td>
<td>0.332</td>
<td>0.309</td>
<td>0.152</td>
<td>0.288</td>
</tr>
<tr>
<td>1988</td>
<td>0.259</td>
<td>0.329</td>
<td>0.294</td>
<td>0.149</td>
<td>0.281</td>
</tr>
<tr>
<td>1990</td>
<td>0.245</td>
<td>0.297</td>
<td>0.270</td>
<td>0.128</td>
<td>0.255</td>
</tr>
<tr>
<td>1991</td>
<td>0.229</td>
<td>0.304</td>
<td>0.269</td>
<td>0.133</td>
<td>0.257</td>
</tr>
<tr>
<td>1992</td>
<td>0.264</td>
<td>0.308</td>
<td>0.273</td>
<td>0.140</td>
<td>0.265</td>
</tr>
<tr>
<td>1993</td>
<td>0.258</td>
<td>0.262</td>
<td>0.260</td>
<td>0.118</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Maximum difference between the upper and lower 95% confidence limits

0.024 0.018 0.018 0.025 0.012

* Proportions; estimated by the Centers for Disease Control and Prevention from the National Health Interview Surveys (2).
approximately one percentage point as compared with the old definition (7). To make the reported estimates of prevalence comparable in our analysis, we subtracted one percentage point from the prevalence data reported in 1992 and 1993 and thus made all the observations consistent with the old definition of prevalence.

Data on age-specific death rates for the general population were obtained for 1970–1990 from the Statistical Abstracts of the United States (8). We assumed 1990 death rates for the years after 1990. Death rates for current smokers were obtained from the 1986 National Health Interview Survey National Mortality Followback Study (9), which estimated age- and gender-specific death rates for current smokers. To combine death rates for males and females into a composite figure for the general population, we weighted the gender-specific death rates by the proportion of male: female current smokers of each age. We used these composite death rates to calculate the age-specific relative risk of death between current smokers and the general population in 1986. To calculate death rates for current smokers in years other than 1986, we multiplied each year’s death rates for the general population by the age-specific relative risk between current smokers and the general population observed in 1986. We acknowledge that using this procedure produced only approximate results. As the prevalence of smoking changes, the death rates for the general population (which includes smokers) will change and so will the relative risk between current smokers and the general population. However, this effect is not large enough to have influenced our results. We performed a sensitivity analysis on this assumption by repeating the analysis described in this paper using just the 1970 death rates for the general population (as opposed to those for 1970–1990); we obtained almost identical results.

Estimation procedure. The dynamic model specified in the previous section was implemented using an EXCEL spreadsheet software program (Microsoft Corporation, Redmond, Washington). We estimated the unknown parameters (the six cessation rates; see expression 6) by selecting the set of values that resulted in the model conforming as closely as possible to past observations on the prevalence of smoking. The method of weighted least squares was chosen as a measure of distance between model response and data. Weights were included in the estimation procedure because the variances in the observed prevalences of smoking among the distinct age groups differed and therefore conveyed different quantities of information for the estimation of parameter values. Observed prevalences were defined as the ratio of the number of smokers in a particular age group to the total number of people in that age group. Because the age groups contained different populations and the sample sizes were not adjusted to reflect this fact, the smaller the population size (and hence the sample size) of the age group, the “noisier” the observed prevalences. The appropriate set of weights to use was one that was proportional to the inverse of the variance of the observations (10). Fortunately, along with the observed prevalences for specific age groups, the data source included the maximum difference between the upper and lower 95 percent confidence limits corresponding to those prevalences (table 1). Assuming that the prevalences were distributed normally for a given sample size, we used this difference to estimate the variance in the observed prevalence. The calculated variances in the observations were used to weight the observations regarding prevalence used in the estimation procedure.

The weighted least-squares estimation procedure was carried out by feeding the response of the dynamic model and the observed data into a minimization routine that controlled the changes in the parameter values until the optimization criterion was satisfied. This combination of dynamic model and optimization routine to estimate parameters has been used previously. For example, Narasimhan et al. (11) used this combination to estimate parameters for a dynamic model of the effects of price and quality on sales response. The optimization routine used to estimate the parameters of our model was the generalized reduced gradient (GRG) algorithm, an optimization technique for solving nonlinear problems. A current version of the algorithm, GRG2, is embedded in the “Solver” component of the Microsoft EXCEL spreadsheet software program. Details on the GRG algorithm are provided by Lasdon et al. (12). Details on the estimation procedure and on other statistical analyses described in this paper are presented in the Appendix.

RESULTS

Estimation results

The estimated values and the lower and upper limits of the 95 percent confidence intervals for the six cessation rates specified in the model are given in table 2. The results show that the smoking cessation rate rose sharply with age and increased for all age groups from 1970 to 1993, consistent with prior expectations. After 1980, coefficients for the rates corresponding to the two older age groups were significantly different from 0. For the period 1970–1980, just the coefficient for the oldest age group was statistically significant. Again, note that particularly during the earlier time

<table>
<thead>
<tr>
<th>Time period</th>
<th>Age group (years)</th>
<th>Estimated value</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower limit</td>
</tr>
<tr>
<td>1970–1980</td>
<td>18–30</td>
<td>-0.09953</td>
<td>-0.02221</td>
</tr>
<tr>
<td></td>
<td>31–50</td>
<td>0.00718</td>
<td>-0.00125</td>
</tr>
<tr>
<td></td>
<td>≥51</td>
<td>0.04528</td>
<td>0.03531</td>
</tr>
<tr>
<td>1981–1993</td>
<td>18–30</td>
<td>0.00209</td>
<td>-0.00689</td>
</tr>
<tr>
<td></td>
<td>31–50</td>
<td>0.02147</td>
<td>0.01586</td>
</tr>
<tr>
<td></td>
<td>≥51</td>
<td>0.05958</td>
<td>0.05441</td>
</tr>
</tbody>
</table>

period, estimates of the cessation rate were confounded by initiation rates. This confounding had the effect of downwardly biasing the estimates and might partially account for the lack of significance in some of the coefficients.

The significance test was conducted by linearizing the model around the estimated values for the parameters and computing their asymptotic covariance matrix. Once the variances of the estimates were calculated, a Z test could be readily performed to test the significance of the coefficients.

The model produced an extremely good fit for the data. The overall adjusted $R^2$ was 0.98. Figure 2 shows the fit of the model to data on the overall prevalence of smoking among adults (all four age groups combined).

Given that the observations for each specific age group constituted a time series of 18 elements, autocorrelation in the series could have biased our estimates of the standard errors of the model parameters. To test for autocorrelation, we first assumed that if it existed, it was defined by an autoregression process of order one (10). Furthermore, we assumed that the same autocorrelation structure existed among the four different age groups. Thus, we merely had to estimate the statistical significance of one correlation coefficient to determine whether autocorrelation existed in the data. Because of gaps in the series, not all the data could be used to determine autocorrelation. After the estimation was completed, and without considering any autocorrelation structure, we computed 1-year differences in the residuals and tested for first-order autocorrelation according to the procedure described by Schmidt (13). The null hypothesis that such autocorrelation did not exist could not be rejected at the 5 percent significance level.

The likelihood ratio test was used to test alternative model specifications. In particular, we tested the hypothesis that no statistically significant differences existed among the post-1980 cessation rates for the three age groups (age 18–30, 31–50, and ≥51 years). We focused on the post-1980 cessation rates because they were used to forecast future prevalence. Explicitly, the following hypotheses were tested:

$$H_0^{(1)}: \nu_4 = \nu_5 \text{ and } H_0^{(2)}: \nu_5 = \nu_6$$

against the following alternative specification:

$$H_a: \nu_4 \neq \nu_5 \neq \nu_6$$

The two hypotheses were rejected at the 5 percent significance level. Thus, we concluded that the three post-1980 estimated cessation rates were significantly different.

Given the fact that the prevalence of smoking has apparently stalled in recent years, it was particularly important to test the hypothesis that cessation rates during the 1990s were not significantly different from those during the 1980s. Thus, we reestimated the model, allowing the cessation rates from 1990 and

FIGURE 2. Observed (•) versus predicted (−) prevalence of smoking among US adults (aged ≥18 years), 1970–1994. Observed prevalence is shown with the 95 percent confidence interval. Data were obtained from the National Health Interview Surveys, 1970–1994 (2,3).
subsequent years to differ from the corresponding rates during the 1980s. Using the likelihood ratio test, we did not find a statistically significant difference between the cessation rates during the 1980s and those during the 1990s at the 5 percent significance level. We concluded that there is no basis for assuming that smoking cessation rates during the 1990s are lower than those that occurred during the 1980s.

**Forecasting results**

Using the post-1980 estimated cessation rates, we forecasted the future prevalence of smoking in the United States until the year 2100 to permit the model to reach a steady-state solution. The size of all future birth cohorts was held constant at the size of the average birth cohort in 1981–1990 (3,643,582). Sensitivity analysis was conducted on the future prevalence of smoking among 18-year-olds. The prevalence of smoking was 25.8 percent among the group aged 18–24 years in 1993. The prevalence among members of this age group has been declining steadily for more than two decades. It has not reached 30 percent since 1985 or 35 percent since about 1975. To forecast future prevalence, we examined the behavior of the model assuming that the prevalence of smoking among 18-year-olds would remain constant at one of the following rates: 20, 25, 30, or 35 percent. At 27 percent, the current prevalence among 18-year-olds is between the two middle figures just cited. The other two prevalences were selected as extreme cases to permit sensitivity analysis. In particular, 35 percent represents a worst-case scenario. Use of a 20 percent prevalence rate among youths reflects the potential impact on the eventual prevalence of smoking among adults of modestly successful policies directed toward reducing the prevalence of smoking among youths.

Figure 3 shows the forecasted prevalence of smoking among US adults given these four selected percentages for the prevalence of smoking at age 18 years. If cessation rates do not decline from those exhibited during the 1980s, the overall prevalence of smoking among US adults will continue to fall, even given an unlikely 35 percent rate for 18-year-olds. The results show that if current cessation behaviors persist, and given a 25 percent prevalence of smoking at age 18 years, the prevalence of smoking among adults will decline from its current level of 25 percent to a steadystate value of 15–16 percent by the second quarter of the next century. Prevalence rates of 20, 30, and 35 percent will produce steady-state prevalence rates among adult smokers of 12.2, 18.4, and 21.5 percent, respectively.

The nearly steady prevalence of smoking among US adults during the 1990s (figure 2) has led many to believe that, short of dramatic new smoking-control policies, prevalence will no longer decline. We examined the increase in the initiation rate (peak prevalence at age 18 years) necessary to stabilize the prevalence of smoking at the current level of 25 percent. We performed this experiment using two sets of cessation rates: those estimated for the post-1980 period and the lower limits of their corresponding 95 percent confidence intervals. Note that the second set of cessation rates corresponds to a very unlikely worst-case scenario. We found that among 18-year-olds, the prevalence necessary to maintain the current level of prevalence among adult smokers (25 percent) was 41 percent for the first set of cessation rates and 35 percent for the second. Current rates of smoking initiation are about 27 percent, which implies that unless initiation rates rise to well above 35 percent, the overall prevalence of smoking among adults will continue.
to decline, regardless of whether significant new tobacco-control initiatives are adopted.

DISCUSSION

Data on the prevalence of smoking during the 1990s suggest that the prevalence among adults has stalled, raising the fear that the historical 25-year decline will not resume. Our results suggest otherwise. If age-specific smoking cessation rates remain stable at the levels exhibited during the 1980s, overall prevalence is likely to continue declining during the remainder of the 1990s and through the next decade. In fact, we predict that the overall prevalence of smoking will continue to fall even if the initiation rate rises to 35 percent, a level not reached since 1975. Given current initiation rates, the decrease in cessation rates necessary to maintain the current prevalence of smoking is extremely unlikely. We are not implying that the prevalence of smoking will drop naturally to 0. On the contrary—we predict that if current conditions persist, the prevalence of smoking will eventually stabilize at 12–22 percent.

Four reasons may account for the fact that the current prevalence of smoking among adults has apparently stalled. First, even though prevalence is still declining (as indicated in figure 2), our model suggests that the rate of decline is slowing. Given this flattening of the prevalence of smoking among adults, it is more difficult than it was 10 years ago to estimate the slope of the series of prevalence observations, for instance, given the same number of data points.

Second, the prevalence of smoking has been sampled far more frequently during the 1990s than it was during any earlier period. An inference of trend based on such closely clustered observations might be clouded by noise in the series.

Third, as discussed previously, in 1992 the National Health Interview Survey changed its definition of current smokers to explicitly include “some days” smokers, many of whom likely escaped categorization as smokers in prior years’ surveys. The Centers for Disease Control and Prevention estimates that this change increased their measure of prevalence by about one percentage point as compared with the old definition. This artificial increase in prevalence creates the impression that when pre- and post-1992 values for the prevalence of smoking are compared, the trend has disappeared. We adjusted their prevalence estimates for 1992 and 1993 downward by one percentage point. The adjustment suggests a continuing decline in prevalence (figure 2).

Fourth, smoking initiation rates among teenagers have risen during the 1990s. As discussed above, these rates are not high enough to stall the decline in the overall prevalence of smoking, but they will definitely slow its decrease, and they provide some explanation for the apparent stabilization of the smoking trend. This surge in smoking among teenagers may reflect a response to significant decreases in the price of cigarettes. For several years now, lower-priced generic and discounted brand cigarettes have become more popular; they now account for over 40 percent of the entire cigarette market. In response to the growing popularity of discounted brands, the major cigarette producers dropped the price of their premium brands by 40 percent in April 1992. The net effect is that the average retail price for cigarettes, adjusted for inflation, is lower today than it was in 1991. A wealth of evidence demonstrates that price significantly affects smoking rates (14).

Our analysis indicates that cessation rates have been reasonably stable for at least 15 years. This finding implies that if the prevalence of smoking is really stalling because of changes in the population’s smoking behavior, this modified behavior is more likely to correspond to initiation rates than to cessation rates. Although we anticipate that the prevalence of smoking among adults will continue to decline even if initiation rates rise to 35 percent, the eventual decline will be modest—to about 21.5 percent. If, in contrast, initiation rates among youths can be reduced to 20 percent from their current level of about 27 percent, adult prevalence will decline to a steady-state rate of just over 12 percent. The difference is substantial in terms of the burden that smoking will place on the health of Americans well into the next century.

This paper attempts to forecast the prevalence of smoking into the next century, assuming that current initiation and cessation trends persist. Through aggressive education and policy efforts, the health toll from tobacco can be reduced to well below its present rate. However, even if our most optimistic assumption about smoking initiation is realized, nearly an eighth of US adults will be smokers in the middle of the next century, and the health toll from smoking will remain extraordinarily high. The health of our children and their children will depend on our not resting on the tobacco-control achievements realized to date.

ACKNOWLEDGMENTS

Preparation of this paper was assisted by a grant from the Robert Wood Johnson Foundation, Princeton, New Jersey. Dr. Warner is principal investigator for the grant.
REFERENCES


APPENDIX

This appendix presents details on the data and analysis mentioned in the “Materials and Methods” section of this paper. Expressions 1–6 in that section outline the forecasting model used.

DATA

Data from the National Health Interview Surveys on the prevalence of smoking among adults were used to estimate parameters (cessation rates) for the model. These prevalences were available for selected years and for the following age groups: age 18–24, 25–44, 45–64, and ≥65 years. These data are shown in table 1 of the text.

PARAMETER ESTIMATION

We used the method of nonlinear weighted least squares to estimate the model. A weighting matrix for the data was used to account for the fact that variances in the prevalence among different age groups were likely to be unequal because the population sizes of the groups differed. The weights were selected to be proportional to the inverse of the data variances so they would closely follow a generalized least squares estimation procedure (10). These variances were estimated from the reported maximum difference between the upper and lower 95 percent confidence limits for the data. If a normal distribution for the data is assumed, this difference is related to the variance by using the following expression:

$$\text{Var} = \left( \frac{\text{difference between the upper and lower 95 percent confidence limits}}{2 \times 1.96} \right)^2$$

The reported upper 95 percent confidence limits and the calculated variances for data in each of the four age groups are shown in appendix table 1.

To implement the method of nonlinear weighted least squares to estimate the model parameters, the following matrices/vectors were identified:

$$\hat{R}_{[52,1]} = (\hat{R}_{(18,24),1970}, \hat{R}_{(25,44),1970}, \ldots, \hat{R}_{(65+,1993)})'$$

observed prevalences

$$\hat{R}_{[52,1]} = (\hat{R}_{(18,24),1970}, \hat{R}_{(25,44),1970}, \ldots, \hat{R}_{(65+,1993)})'$$

calculated prevalences

$$\Omega_{d(4,4)} = \begin{bmatrix} 0.3748 & 0.2109 \\ 0.2109 & 0.4067 \end{bmatrix}$$

1-year weighting matrix

APPENDIX TABLE 1. Calculated variances for data on the prevalence of smoking, by four age groups of US adults, 1970–1993

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>18–24</th>
<th>25–44</th>
<th>45–64</th>
<th>≥65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum difference between upper and lower 95% confidence limits (%)</td>
<td>2.4</td>
<td>1.8</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Variance</td>
<td>0.3748</td>
<td>0.2109</td>
<td>0.2109</td>
<td>0.4067</td>
</tr>
</tbody>
</table>
The vector of cessation rates was estimated by solving the following optimization problem:

$$\text{Min} \left( \frac{1}{N} \sum_{i=1}^{N} \left[ \tilde{R} - \hat{R}(N) \right]^{'} \Omega^{-1} \left[ \tilde{R} - \hat{R}(N) \right] \right)$$

This minimization problem was solved using the generalized reduced gradient (GRG) algorithm, embedded in the "Solver" component of the Microsoft EXCEL spreadsheet software program. The estimated values for the cessation rates are given in table 2 of the text.

**CONFIDENCE INTERVALS**

Ninety-five percent confidence intervals were developed for the estimated cessation rates. The asymptotic covariance matrix of these estimated coefficients is as follows (10):

$$\delta^2 \left( \frac{\partial \hat{R}(N)}{\partial N} \right)' \Omega^{-1} \left( \frac{\partial \hat{R}(N)}{\partial N} \right)$$

where the multiplicative scalar in the expression can be computed as follows:

$$\delta^2 = \frac{\left[ \tilde{R} - \hat{R}(\hat{N}) \right]' \Omega^{-1} \left[ \tilde{R} - \hat{R}(\hat{N}) \right]}{(52 - 6)}$$

Partial derivatives in the above expression were computed numerically using an EXCEL spreadsheet. Matrix computations were performed using MATLAB software (The MathWorks, Inc., Natick, Massachusetts). The resulting asymptotic covariance matrix for the estimated cessation rates follows:

\[
\begin{bmatrix}
0.001479 & 0.0005 - 0.0012 - 0.0032 - 0.0060 & 0.0008 \\
0.0005 & 0.0125 & 0.0013 & 0.0019 & - 0.0040 & - 0.0023 \\
- 0.0012 & 0.0013 & 0.0175 & - 0.0002 & 0.0009 & - 0.0062 \\
- 0.0032 & 0.0019 & - 0.0002 & 0.0142 & - 0.0016 & 0.0000 \\
- 0.0060 & - 0.0040 & 0.0009 & - 0.0016 & 0.0055 & - 0.0003 \\
0.0008 & - 0.0023 & - 0.0062 & 0.0000 & - 0.0003 & 0.0047
\end{bmatrix}
\]

The 95 percent confidence intervals for the cessation rates were obtained by assuming that these parameters were asymptotically normally distributed. For instance, the 95 percent confidence interval for \( \nu_1 \) is as follows:

$$-0.00953 \pm 1.96 \times \sqrt{0.001479 \times 0.0283}$$

The lower and upper limits of the 95 percent confidence intervals for the estimated cessation rates are shown in table 2 of the text.

**ALTERNATIVE MODELS**

Several alternative specifications for the cessation rates were formulated to test whether the data supported a more parsimonious model. Since the post-1980 cessation rates were to be used for forecasting, restrictions were imposed on such rates in the alternative specifications to simplify the model. Specifically, the following alternative models, identified by the restrictions on the \( \nu \)'s, were proposed:

\[H_0^{(1)}: \nu_4 = \nu_5\]
\[H_0^{(2)}: \nu_5 = \nu_6\]

These two models were tested against the alternative specification:

\[H_0: \nu_4 \neq \nu_5 \neq \nu_6\]

which corresponds to the original (unrestricted) model discussed in the paper.

We tested the proposed restrictions on the model using the likelihood ratio test (10). The following notation describes the procedure:

$$S(N) = \left[ \tilde{R} - \hat{R}(\hat{N}) \right]' \Omega^{-1} \left[ \tilde{R} - \hat{R}(\hat{N}) \right]$$
APPENDIX TABLE 2. Test of alternative models for determining the rates of smoking cessation among US adults, 1970–1993

<table>
<thead>
<tr>
<th>Model (i)</th>
<th>$\lambda_{LR}^{(i)}$</th>
<th>$J$</th>
<th>$\chi^2(4,0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.61</td>
<td>1</td>
<td>3.84</td>
</tr>
<tr>
<td>2</td>
<td>21.16</td>
<td>1</td>
<td>3.84</td>
</tr>
</tbody>
</table>

$\bar{N}^{(i)}$ = cessation rates estimated under model restriction (i)

$\bar{N}$ = as before, cessation rates estimated under the “full” model

$\lambda_{LR}^{(i)} = 52\{\ln[S(\bar{N}^{(i)})] - \ln[S(\bar{N})]\}$

At the 5 percent significance level, using the likelihood ratio test calls for rejection of the proposed restricted model (i) in favor of the full model if $\lambda_{LR}^{(i)}$ exceeds $\chi^2(J,0.05)$, where the number of degrees of freedom $J$ is equal to the number of restrictions that the proposed model (i) imposes on the full model. Appendix table 2 shows, for each proposed model, its calculated $\lambda_{LR}^{(i)}$ and the critical value necessary to reject the restrictions. The values in this table indicate that the two restricted models were rejected in favor of the unrestricted model.

We used the likelihood ratio test to examine the hypothesis that 1990s cessation rates do not differ from 1980s rates. To test this hypothesis, we reestimated the model allowing cessation rates from 1990 and subsequent years to differ from those that occurred during the 1980s. Thus, we added three new parameters to the model: $v_1$, $v_2$, and $v_3$, which represent the 1990s cessation rates corresponding to the age groups 18–30, 31–50, and >50 years, respectively. Specifically, we tested the following restrictions:

$H_0: v_1 = v_2 = v_3$ versus a fully unrestricted model, where all cessation rates are allowed to differ.

For this test, we obtained $\lambda_{LR} = 2.60$. With three restrictions, at the 5 percent significance level, the critical value needed to reject the null hypothesis was $\chi^2(3,0.05) = 7.81$. Therefore, we cannot reject the restricted model and conclude that the 1990s cessation rates are no different from the 1980s rates. In fact, the $p$ value for this test is 30–50 percent.

AUTOCORRELATION

First-order autocorrelation in the model was tested according to the procedure described by Schmidt (13). Appendix table 3 gives the residuals ($e_i$) of the model (observed minus predicted value) that correspond to contiguous observations.

Under the null hypothesis that the first-order autocorrelation in the residuals is equal to 0, the statistic

$$\sqrt{\frac{N}{56}} \times \frac{\sum(e_i \times e_{i-1})}{\sum e_i^2}$$

is asymptotically normally distributed with a mean of 0 and a variance of 1 (13). At the 5 percent significance level, we would reject the null hypothesis if the absolute value of the statistic were greater than 1.96. When evaluating the statistic given our data, we obtained the following:

$$\sqrt{\frac{56}{56}} \times \frac{0.000508}{0.002556} = 1.19 < 1.96$$

Therefore, we do not reject the null hypothesis that there is no first-order autocorrelation in the data, and our estimates of the standard errors are appropriate.

APPENDIX TABLE 3. Model residuals* for contiguous observations on the prevalence of smoking among four age groups of US adults

<table>
<thead>
<tr>
<th>Year</th>
<th>Age group (years)</th>
<th>18-24</th>
<th>25-44</th>
<th>45-64</th>
<th>≥65</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td></td>
<td>-0.00514</td>
<td>-0.00266</td>
<td>0.00132</td>
<td>0.00146</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td>-0.00237</td>
<td>-0.00160</td>
<td>-0.01373</td>
<td>0.00859</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td>-0.00761</td>
<td>-0.00850</td>
<td>-0.00135</td>
<td>0.02198</td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td>-0.00906</td>
<td>0.00243</td>
<td>0.00992</td>
<td>0.00214</td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td>-0.00905</td>
<td>0.00667</td>
<td>0.00309</td>
<td>0.00250</td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td>0.00116</td>
<td>-0.01214</td>
<td>-0.00628</td>
<td>-0.01028</td>
</tr>
<tr>
<td>1991</td>
<td></td>
<td>-0.00385</td>
<td>-0.00018</td>
<td>-0.00096</td>
<td>-0.00073</td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td>0.01687</td>
<td>0.00458</td>
<td>0.00386</td>
<td>-0.00383</td>
</tr>
<tr>
<td>1993</td>
<td></td>
<td>0.00644</td>
<td>-0.00357</td>
<td>-0.00159</td>
<td>-0.02446</td>
</tr>
</tbody>
</table>

* Proportions.